Evolutionary History of the Angiosperms and Its Relevance to Brassica

A.H. Paterson[1], J.E. Bowers[1], J.C. Estill[1], Thomas C. Osborn[2], J.C. Pires, R. Amasino[4], C.F. Quiros[5] and M. Farnham[6]
[1] Plant Genome Mapping Laboratory, University of Georgia, Athens, GA, USA
[2] Department of Agronomy, University of Wisconsin, Madison, WI, USA
[3] Department of Biological Sciences, University of Missouri, Columbia, MO, USA
[4] Department of Biochemistry, University of Wisconsin, Madison, WI, USA
[5] Department of Vegetable Crops, University of California, Davis, CA, USA
[6] US Vegetable Laboratory, USDA-Agricultural Research Service, Charleston, SC, USA

Keywords: *Arabidopsis*, paleopolyploid, comparative genomics, phylogenomics, physical map, bacterial artificial chromosome

Abstract

One of the more surprising discoveries in the sequences of *Arabidopsis* and rice has been that both are ancient polyploids, in the case of *Arabidopsis* involving at least two and perhaps three duplications of most of the genome. Ancient whole-genome duplication and subsequent loss of many of the duplicated gene copies appears to contribute substantially to deviations from co-linearity and/or synteny that are observed in many angiosperm lineages. The *Brassica* genomes are shaped not only by these ancient polyploidization events shared with *Arabidopsis*, but by 1-2 additional events in the 'diploid' *Brassicas* and the recent formation of new polyploid species. An important dimension in clarifying the evolutionary history of the *Brassica* genus is the better resolution of the timing and consequences of polyploidizations since the *Brassica-Arabidopsis* divergence. Early efforts to assess the predictive value of *Arabidopsis* data for determining fates of duplicated *Brassica* genes will be discussed.

INTRODUCTION

Domestication of most crops resulted in enhancement of a single plant part for use by humans, such as the seeds of seed/grain crops, the fruits of some trees, or the roots of some vegetable crops. The *Brassica* crops are a notable exception, including forms that have been selected for enlarged vegetative meristems at the apex (cabbages) or in the leaf axils (Brussels sprouts), forms with a proliferation of floral meristems (broccoli) or aborted floral meristems (cauliflower), forms with swollen stems (kohlrabi, marrowstem kale), or orate leaf patterns (kales). This wide range of variation is especially surprising in that it is all found in freely-intercrossing members of a single species, *Brassica oleracea*. Additional variation is found in close relatives such as the enlarged and twisted leaves of Pak-choi (*B. rapa* subsp. *chinensis*) and Chinese cabbage (*B. rapa* subsp. *pekinensis*); enlarged roots of turnip (*B. rapa* subsp. *rapifera*); and high seed yields of oilseed *B. rapa* and *B. napus*.

Detailed elucidation of comparative genome structure and function of the three cultivated diploid *Brassicas* and morphotypes within these species, is an exciting opportunity to leverage the *Arabidopsis* sequence into new knowledge about the functions of specific genes affecting many agriculturally-important aspects of plant growth and development, and new understanding of the causes of biological diversity. *Brassica* and *Arabidopsis* are thought to have shared common ancestry only 15-20 million years ago (Yang et al., 1999), a much closer relationship than that among grasses such as rice and maize that are often thought of as primary beneficiaries of comparative genomic approaches (Paterson et al., 2001).

The ability to cross divergent morphotypes to obtain progeny with normal transmission genetics adds an important dimension to our work, by using QTL mapping to determine numbers and locations of loci that control morphological divergence. While a few genes like the homologs of *Arabidopsis* mutants such as "*CAULIFLOWER*" play

major roles in some *Brassica* morphologies (Kempin et al., 1995; Smith and King, 2000; Purugganan et al., 2000), it has long been known that these morphologies are under complex genetic control (e.g., Pease, 1926; Detjen et al., 1933; Currence, 1934; Yeager, 1943). Some *Brassica* QTL map to locations that correspond to known *Arabidopsis* mutants - but many do not (e.g., Lan and Paterson, 2000b, 2001). While *Arabidopsis* information and tools are of high value for *Brassica* genomics, to think that we could rely largely on *Arabidopsis* data to identify key *Brassica* genes would be to disregard the 15-20 million years of evolution that account for the remarkable diversity among and within *Brassica* species.

RESULT AND DISCUSSION

The recent confirmation (Blanc et al., 2000; Paterson et al., 2000; AGI, 2000; Vision et al., 2000) of early hints (Kowalski et al., 1994) that *Arabidopsis* is an ancient polyploid has necessitated re-thinking the comparative approaches by which we approach the 'translation' of its genomic data to utilization in major crops. Crucial to appropriate genomic comparison is knowledge of whether any polyploidization events occurred before, or after, taxonomic divergence (Kellogg, 2003); we have described methods to determine this (Bowers et al., 2003; Chapman et al., 2004).

Brassica researchers are somewhat more fortunate than those in other taxa seeking to benefit from the *Arabidopsis* sequence, in that all three of the large-scale *Arabidopsis* duplication events (Bowers et al., 2003) occurred well before the *Brassica-Arabidopsis* divergence. Consequently, much of the 'diploidization' that has taken place since each of these three events is shared by *Brassica*, simplifying to some degree the alignments of the respective genomes and cross-utilization of genomic data.

However, there remain considerable gaps in our knowledge of the diversification of the *Brassica* lineage subsequently to its divergence from *Arabidopsis*. Of singular importance will be to investigate the consequences of large-scale, probably genome-wide duplication(s) that occurred in the *Brassica* lineage (Fig. 1). That one or more such events has happened is irrefutable, based on many independent discoveries of large duplicated segments in the genetic maps of each of the three cultivated diploids [*B. rapa* (syn. *rapa*, 2n=20, genome AA), *B. nigra* (2n=16, genome BB) and *B. oleracea* (2n=18, genome CC)]. While the ubiquity of the event(s) implies that genomic duplication may have preceded the divergence of the cultivated diploid *Brassicas* from one another, this is by no means proven. In particular, the considerable divergence of *B. nigra* from *rapa/oleracea* raises the possibility of independent duplications.

Further, at present we can only speculate about the exact antiquity of the duplication event(s), and have only rough estimates of the timing of divergence among the cultivated diploid *Brassicas* (e.g. Wroblewski et al., 2000). Consequently, the relative duration for which the diploid *Brassicas* evolved together or independently after polyploid formation is unknown. In view of both rapid restructuring (e.g. Song et al., 1995; Pires et al., 2004a) and long-term diploidization processes (e.g. Bowers et al., 2003) that affect polyploid genomes, answers to these questions are essential to making sense of the comparative genome structure and function of the diploid *Brassicas*. If species divergence closely followed polyploidization, then most restructuring may have occurred independently in the three lineages, contributing to variation among the species. Virtually all gene probes hybridize to genomic DNA from each species, but we lack information about how frequently the species vary in copy number for particular genes.

The need for clarifying the timing of the genome structural changes in the lineage is complicated by the fact that the phylogeny of *Brassica* species is not straightforward because the genus *Brassica* (ca. 40 species) is not a monophyletic group and is intermingled with *Diplotaxis*, *Erucastrum*, *Moricandia*, *Raphanus*, *Sinapis* and other genera. While the tribe Brassiceae (ca. 50 genera, 240 species) is a well-defined monophyletic group based on morphological and molecular data, several subtribes and genera are in need of phylogenetic and taxonomic revision (Anderson and Warwick, 1999; Warwick and Black, 1991, 1993, 1997; Appel and Al-Shebaz, 2002; Koch et al., 2003).

The promise of a finished sequence of *Brassica rapa* opens new doors into a wide range of questions regarding the structure, function and evolution of the *Brassica* genomes, but represents only a beginning. What is really needed to realize the potential learning opportunities that lie hidden in the *Brassica* genomes, and also provide intrinsic low-cost genetic solutions to the challenges that face *Brassica* crop producers worldwide, is representative sequences from each of the species in the genus. Technology will eventually bring this goal within reach, very probably within our generation. However, until the cost of sequencing the entire genome drops to a nominal level (and even then), genetically-anchored physical maps will be an important tool in the study of genomic diversity. Recent efforts in several countries to build and characterize bacterial artificial chromosome (BAC) libraries for a diverse sampling of *Brassica* species and morphotypes are of high value, and can be further enhanced by integration based on hybridization to large numbers of common DNA probes. In the short term, these physical maps will provide for localized genomic comparisons and the isolation of morphotype-specific alleles for a wide range of studies. In the long view, these maps may provide a foundation for sequencing of additional *Brassica* species and morphotypes.

Genomic studies of other groups in the Brassicales order offer many additional learning opportunities. For example, physical mapping of a member of the Caricaceae, *Carica papaya* (Liu et al., 2004) promises new insights into the genetics of sex determination in plants. More generally, this work may shed light on the early stages of sex chromosome evolution that are not accessible in well-studied systems such as mammals (in which sex chromosomes are ancient). The Brassicales are likely to hold many other novelties to be uncovered by energetic and creative investigations. Our generation of scientists is indeed fortunate to be positioned to unravel the multi-million year histories of our subject taxa at the DNA level, and to reveal how these histories have influenced botanical diversity and agricultural productivity.

ACKNOWLEDGEMENT

We (R.A., T.O., A.H.P., C.F.Q.) thank the US Department of Agriculture for financial support of our work.

Literature Cited

Anderson, J.K. and Warwick, S.I. 1999. Chromosome number evolution in the tribe *Brassiceae* (Brassicaceae): evidence from isozyme number. Plant Systematics and Evolution 215:255-285.

Appel, O. and Al-Shebaz, I.A. 2002. Cruciferae. p.75-174. In: K. Kubitzki (ed.), Families and genera of vascular plants, vol. 5, Spinger-Verlag, Berlin, Heidelberg.

Arabidopsis Genome Initiative. 2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408:796-815.

Bowers, J.E., Chapman, B.A., Rong, J. and Paterson, A.H. 2003. Unravelling angiosperm chromosome evolution by phylogenetic analysis of chromosomal duplication events. Nature 422:433-438.

Chapman, B.A., Bowers, J.E., Schulze, S.R. and Paterson, A.H. 2004. A comparative phylogenetic approach for dating whole-genome duplication events. Bioinformatics 20:180-185.

Currence, T.M. 1934. Results from hybridizing cabbage with Brussels sprout. Proc. Am. Soc. Hort. Sci. 32:485-487.

Detjen, L.R. and McCue, C.A. 1933. Cabbage characters and their heredity. Delaware AES Tech Bull. 180:1-127.

Kellogg, E.A. 2003. It's all relative. Nature 422:383-384.

Kempin, S.A., Savidge, B. and Yanofsky, M.F. 1995. Molecular basis of the cauliflower phenotype in Arabidopsis. Science 267:522-555.

Koch, M., Haubold, B. and Mitchell-Olds, T. 2001. Molecular systematics of the Brassicaceae: Evidence from coding plastidic *mat*K and nuclear *chs* sequences. Amer. J. Bot. 88:534-544.

Kowalski, S.D., Lan, T.-H., Feldmann, K.A. and Paterson, A.H. 1994. Comparative mapping of *Arabidopsis thaliana* and *Brassica oleracea* chromosomes reveals islands of conserved gene order. Genetics 138:499-510.

Lan, T.H., Delmonte, T.A., Reischmann, K.P., Hyman, J., Kowalski, S., McFerson, J., Kresovich, S. and Paterson, A.H. 2000a. EST-enriched comparative map of *Brassica oleracea* and *Arabidopsis thaliana*. Genome Research 10:776-788.

Lan, T.H. and Paterson, A.H. 2000b. Comparative evolution of QTLs sculpting the curd of *Brassica oleracea*. Genetics 155:1927-1954.

Lan, T.H. and Paterson, A.H. 2001. Comparative mapping of QTLs determining the plant size of *Brassica oleracea*. TAG 103:383-397.

Liu, Z., Moore, P.H., Ma, H., Ackerman, C.M., Ragiba, M., Pearl, H.M., Kim, M.S., Charlton, J.W., Yu, Q., Stiles, J.I., Zee, F.T., Paterson, A.H. and Ming, R. 2004. A Primitive Y chromosome in papaya marks incipient sex chromosome evolution. Nature 427:348-352.

Paterson, A.H. et al. 2000. Comparative genomics of plant chromosomes. The Plant Cell 12:1523-1539.

Paterson, A.H., Lan, T.H., Amasino, R., Osborn, T.C. and Quiros, C.F. 2001. Brassica genomics – a complement to, and early beneficiary of, the *Arabidopsis* sequence. Genome Biol. 2:1011.1-1011.4

Pease, M.S. 1926. Genetic studies in *Brassica oleracea*. J. Genet. 16:363.

Pires, J.C., Zhao, J., Schranz, M.E., Leon, E.J., Quijada, P.A., Lukens, L.N. and Osborn, T.C. 2004a. Flowering time divergence and genomic rearrangements in resynthesized Brassica polyploids Brassicaceae. Biol. J. Linnaen Soc. 82:675-688.

Purugganan, M.D., Boyles, A.L. and Suddith, J.I. 2000. Variation and selection at the CAULIFLOWER floral homeotic gene accompanying the evolution of domesticated *Brassica oleracea*. Genetics 155:855-862.

Smith, L.B. and King, G.J. 2000. The distribution of BoCAL-a alleles in *Brassica oleracea* is consistent with a genetic model for curd development and domestication of the cauliflower. Molec. Breed. 6:603-613.

Vision, T.D., Brown, D.B. and Tanksley, S.D. 2000. The origins of genomic duplications in *Arabidopsis*. Science 290:2114-2117.

Warwick, S.I. and Black, L.D. 1991. Molecular systematics of Brassica and allied genera (subtribe Brassicinae, Brassiceae)-chloroplast genome and cytodeme congruence. Theor. Appl. Genetics 82:81-92.

Warwick, S.I. and Black, L.D. 1993. Molecular relationships in subtribe Brassicineae (Cruciferae, tribe Brassiceae). Canad. J. Bot. 71:906-918.

Warwick, S.I. and Black, L.D. 1997. Phylogenetic implications of chloroplast DNA restriction site variation in subtribes Raphaninae and Cakilinae (Brassicaceae, tribe Brassiceae). Canad. J. Bot. 75:960-973.

Wroblewski, T., Coulibaly, S., Sadowski, J. and Quiros, C.F. 2000. Variation and phylogenetic utility of the *Arabidopsis thaliana* Rps2 homolog in various species of the tribe Brassiceae. Molecular Phylogenetics and Evolution 16:440-448.

Yang, Y.W., Lai, K.N., Tai, P.Y. and Li, W.H. 1999. Rates of nucleotide divergence in angiosperm mitochondrial DNA sequences and dates of divergence between Brassica and other angiosperm lineages. J. Mol. Evol. 48:597-604.

Yeager, A.E. 1943. The characteristics of crosses between botanical varieties of cabbage (*Brassica oleracea*). Proc. Am. Soc. Hort. Sci. 43:199.

Figures

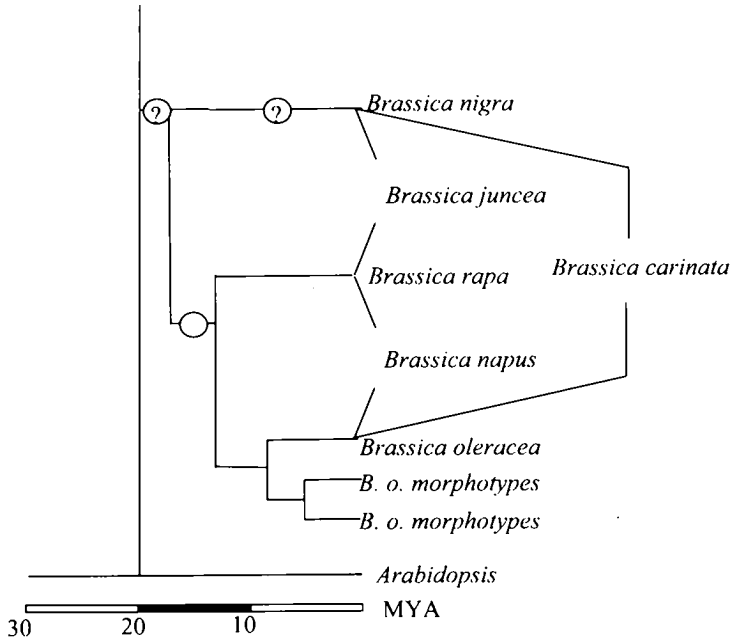

Fig. 1. Superimposing the history of genomic duplications on the *Brassica* family tree. While it is widely accepted that there have been at least one, perhaps two large-scale chromosomal duplications in the *Brassica* lineage since its divergence from *Arabidopsis*, it remains uncertain when exactly these events have occurred in relation to the divergence of the diploid *Brassica* species from one another. Similarity in their genetic maps suggests that most if not all of these events predated the divergence of *B. rapa* and *B. oleracea*, however whether *B. nigra* shares one or both of these duplications or has undergone independent duplications remains unknown. More information about the relative duration for which the diploid *Brassicas* evolved together or independently after polyploid formation is important to better understanding of their comparative genomics and relating genomic differences to functional/morphological diversity in the genus.